Professur Multimediatechnik, Privat-Dozentur Angewandte Informatik

Prof. Dr.-Ing. Klaus Meißner

PD Dr.-Ing. habil. Martin Engelien

(Hrsg.)

GeNeMe '06

Gemeinschaften in Neuen Medien

an der
Fakultät Informatik der Technischen Universität Dresden

unter Mitwirkung des
Bundesministeriums für Bildung und Forschung,
Programm Innovative Arbeitsgestaltung und der
Gesellschaft für Informatik e.V.
GI-Regionalgruppe Dresden

am 28. und 29. September 2006 in Dresden

http://www-mmt.inf.tu-dresden.de/geneme2006/

geneme@mail-mmt.inf.tu-dresden.de

D.7 Towards a Social Network Intelligence Tool for visual Analysis of Virtual Communication Networks

Matthias Trier

Technical University Berlin, Institute for Business Informatics

Abstract

Communities of Practice regularly utilize virtual means of communication. The according software support provides its members with many sophisticated features for generating content and for communicating with each other via the internet or intranet. However, functionalities to monitor, assess, coordinate, and communicate the quality and development of the underlying electronic networks of experts are frequently missing. To meet this need of increased manageability, this contribution introduces a Social Network Intelligence software approach which aims at supporting the comprehension of the structure and value of electronic communities by automatically extracting and mining available electronic data of various types of virtual communication networks, like e-mail archives, discussion groups, or instant messaging communication. Experimental structural visualizations employing Social Network Analysis methods are combined with Keyword Extraction to move towards a Social Network Intelligence approach which generates transparency of complex virtual communication networks. Together with a comprehensive visualization method, an approach for software-supported communication network measurement and evaluation is suggested. It supports the identification of important participants, topics, or clusters in the network, evaluates the interpersonal communication structure and visually traces the evolvement of the knowledge exchange over time.

1. Introduction

In the past years, the event of sophisticated means of global computer-mediated communication (CMC) has facilitated the development of Communities of Practice (CoPs) whose members are not co-located [LeSt01]. This draws the attention towards the issue of virtual community. Rheingold [Rhei93] defines such virtual groups as democratic and equal coalitions of individuals, which effectively cooperate in a joint venture.

Despite the virtuality and the large size of electronic groups, a survey of Berge and Collins [BeCo00] substantiates that there is a perception of community. The authors identified, that 72.9 percent of virtual networks' moderators considered their group as a community. Further, 70.6 percent of the moderators believed that their members feel themselves as part of a community. More than 70 percent where also noting, that they actively promote the sense of being a community.

The increased practical employment of virtual communication networks for corporate knowledge intensive work poses new challenges. The abundance of communication channels, the very large group size, the underlying voluntary participation with the resulting absence of the ability to execute hierarchical authority, and the complex issue of recognizing and utilizing value creation results in difficulties in understanding and utilizing this complex organizational structure: A recent study of Ambrozek and Cothrel [AmCo04] shows, that although 79 percent of moderators and members of CoPs agree, that technologies for online communities are continuing to improve and participation in online communities is growing (82 percent), most organizations can't measure return on investment (72 percent agree). To a large extend, this is attributable to the fact, that the discipline of creating and managing communities is poorly defined (59 percent agree). This results in a situation, where less than half of the respondents feel, that executives understand the value of online communities.

Similar insights yielded the analysis of communities conducted by the American Productivity and Quality Center APQC [APQC03]. It identified systematic monitoring of effectiveness and assessing the 'health' of the community as being a very important factor for knowledge management in an enterprise. Here, besides the incorporation of general strategic objectives of the organization and leadership qualifications of the moderating persons, the community structure is named an important element of management. This institution further emphasized performance measurement using monitoring and controlling instruments.

These findings contrast the currently offered software functionality for virtual groups. It primarily targets the user group of members and does not supply sophisticated functionality which supports transparency, coordination, monitoring, or management. Further, the applications tend to concentrate on content oriented features. The very important social domain and also underlying processes of management are, despite some exceptions, largely ignored. Transparency in this domain would provide a better impression of identity, prominence, and social mutuality in a network. In this context, Erickson and Kellogg [ErKe00] express the necessity of providing visible clues based on perceptual information of the social situation (presence and activities of users) to the members of a virtual group in order to create social resources that help to structure the online interactions. By such visual information, people become aware of each other and social conventions and dynamics are enriched. The mutual awareness increases the accountability of the member's actions. Behavior is more visible and persistent, the history and thus also the character of users is getting conveyed. This creates more coherent, productive, and fluid online interactions. Users can imitate and observe others, peer pressure emerges.

The simple logging facilities provided by current software are not targeted at such network-oriented collaborative work and ignore the benefits of network analysis. Most of the required data for a comprehensive and methodological measurement (like logins, contributions, and references) is usually electronically archived and can be utilized for analysis. Additionally to this gap in software applications, companies are often only conducting manual survey-based audits to assess their communities, ignoring the rich data they could derive from their software and ignoring network metrics and models to improve the effectiveness of a community.

This situation implies two main research issues:

1. What data offered by available communication means of CoP-supporting software provides most value for analyzing, visualizing, and developing the virtual group?
2. How can virtual communication networks be modeled, analyzed, and visualized by software based Social Network Intelligence methods to enhance transparency for moderators, analysts, and members of communities?

2. Research Objective and Methodology

Taking the research issues of the previous section into account, this paper introduces a software based method and the related tool Commetrix which aims at generating visual insights into structures of existing virtual communication networks [Trie05a, Trie05b].

Utilizing the available electronic data of virtual communication networks, the objective of this research is to apply information systems to discover active or inactive areas of networking (e.g. via discussion) as well as their defining properties. This helps members, moderators, and researchers to understand community structures better. The software design focus is on simple and automated data import from a wide variety of electronic sources, insightful visualizations or animations, and a systematic approach for measurement and evaluation.

The research methodology included extensive literature reviews to derive preliminary requirements for members, researchers, and moderators, observation of running knowledge exchange in virtual communities, further exploratory interviews with moderators, iterative prototypical software engineering to translate the identified requirements into preliminary visualizations and functionality, and subsequent field testing using public virtual and corporate community networks to determine useful measures and to refine the technical approaches. The developed system aims at providing an external addition (‘add-on’) to current systems for communication and collaboration support (e.g. discussion boards, groupware, e-mail etc.).

The next chapters will show the underlying modeling approach to subsequently introduce the developed software functionality for modeling, visualizing, and analyzing virtual communities to support CoP monitoring and facilitation.

3. Towards a modeling perspective for virtual Communities

Communities of Practice utilize the special organizational form of a network to enable the flow of information and influence from the top to the bottom and vice versa, but also horizontally through the relations [StLi00]. With these properties, a Community of Practice is a flexible knowledge network that overcomes existing rigid hierarchical communication structures as it allows for virtually unlimited configurations. Barley [Barl96] comes to a similar conclusion: If in an organizational structure, knowledge and capabilities were generated domain-specific, the work is less determined by the vertical chain of command (influence) but rather by (lateral or) horizontal communication and collaboration between different groups. All these contributions imply that the community provides its value through its network. The network delivers fix points for the combination of order and chaos in the organization. Every employee must be enabled to purposefully switch between the processing of expected and unexpected information. The necessary orientation is provided by a network, which is offering active connections (for coordination and communication), but even more importantly connections, which potentially can be activated on demand in special situations [Baec99, p.26]. In this context, Lesser and Storck [LeSt01] emphasize, that "communities play a significant role in the development of social capital, which in turn influences organizational outcomes". This social capital is indicating the resources one can access by utilizing established social relations [cf. NaGo98]. Hence, modeling virtual communities from a people network perspective showing their relationships and expertise should allow for understanding the activities, structures, and value generation of Communities of Practice. This is substantiated by extensive empirical research on social networks. It confirms that interactions of people captured over time form a network structure via communication [Krac91]. On a more detailed level of analysis, this branch of research also found, that a network is usually not homogeneous, but has structural properties, like clusters and structural holes. The configurations of the actors' relations differ and so do their roles in the network. A final statement which highlights the importance of modeling a Community of Practice as a collaborative network of experts comes from Swan [Swan01], whose "research indicates" that: "It seems more likely that the key to achieving coordinated action does not so much depend on those 'higher-up' collecting more and more knowledge as on those 'lower-down' finding more and more ways of getting connected and interrelating the knowledge each one has." [Swan01, p.8].

These insights motivated the development of a model to visualize community structures and processes, which is based on an integration of graph theory, social network analysis, and text mining methods (cf. Figure 2). The data model includes the elements author and relation, together with a wide array of modeled attributes and properties. The latter additionally convey a qualitative or quantitative meaning, which can be grouped, compared or measured.

Each author is represented by a node (sphere). This node has core attributes like name, e-mail, and index number. Author properties can be organizational affiliation, number of contacts, or organizational hierarchy. Relations represent an additive set of messages and are represented by edges. They can have properties, like the number of messages or the average evaluation of their contents. To convey additional information clues about relevant network properties, the representation of author nodes and their relationship network can be extended by text labels, different node sizes, rings around nodes, or node colors. Further insights are provided by integrating elicited keyword tags showing the contents of the network's information exchange. Finally, longitudinal animations and measures of the communication network model help to identify and communicate evolution and growth [also cf. Trie05a].

4. Introducing automated elicitation of community models

In order to compute and present the visual model of the analyzed communication network, the Commetrix software connects to various communication platforms. Here, an extendable set of data extraction connectors has been developed. A connector's objective is to elicit useful data about the information exchange between participants of some online group. Currently, the solution allows for analyzing e-mail archives, instant messaging protocols, newsgroup archives, and Slashdot discussions.

One fundamental issue in automatically capturing data using a source-independent approach results from two possible storage paradigms for electronic discourses: peer-oriented networks versus hierarchical networks. Whereas instant messaging, e-mail, or chats belong to the first category, discussion groups include hierarchical information, as messages can refer to parent messages.

In Figure 1, the two different storage paradigms are compared. In peer-oriented archives, author A is contacting author B directly via the message. In hierarchical archives, author A is writing message 1 which references message 2, which has been written by author 2.

After preparing the data, the community network model is generated. The layout component is a central feature and applies the Fruchterman and Reingold Spring Embedder Algorithm [FrRe91]. Using this algorithm the people network can be presented in meaningful clusters where actors with strong links are located close to each other. The developed layout component generates two dimensional and three dimensional layouts for the presentation of the communication network model. The third dimension leaves more room for the nodes to arrange themselves, thus creating a better image of the network structure, which often resembles chemical molecules.

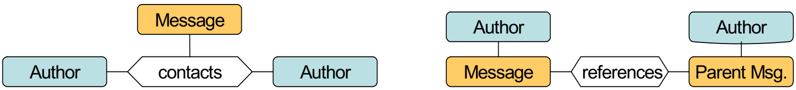

Figure 1: Integrating different storing paradigms in electronic discourses

The complexity of the resulting visual graph model can be reduced via a configurable set of filters. e.g. time filtering. It shows only the nodes which are active in a given time period. This necessitates an extension of the Fruchterman and Reingold algorithm to allow for longitudinal analysis and animation of the evolvement of the networks (see next sections).

5. Analyzing sample communication networks

To illustrate the improved transparency and the resulting benefit for understanding virtual communication structures by visualizing communication networks using the Commetrix model, various example outputs will be briefly demonstrated now. A first (public) sample is a dataset of a ‘Slashdot’ discussion. The communication between authors is collected via automatically connecting to the website that hosts the conversation. The 'Slashdot' discourse employs a hierarchical storage paradigm (cf. Figure 1). Its individual properties are identified and modeled, e.g. the evaluation of authors. From this data the author network model shown in figure 2 is being generated. The distance between authors is smaller, the stronger their communication relation is, i.e. the more they have communicated with each other (also represented by edge thickness). The node size has been set to represent the activity of the authors, the node labels show the authors’ names, the edge labels show the amount of communication between two nodes (as this is only a small sample of a short period, there has not been much communication between any two authors). The node color has been set to represent the author evaluation. These model settings yield the first insights: The most active authors do not need to be the ones with the best evaluation; rather they might be littering the network. The node size could also be set to represent the average amount of communication with any of the author's contacts, or in other words the average depth of his relationships. This uncovers, if the authors maintain many weak relationships or rather only few but strong relationships. Alternatively, setting the node size to represent messages received would yield in insights about the prominence of authors. The nodes who received the most attention ('prominence') from others are bigger in size.

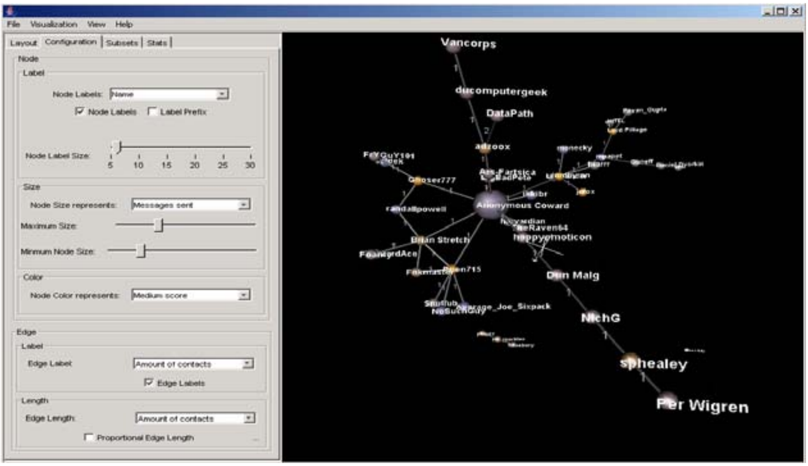

Figure 2: A sample communication network shown in the Commetrix GUI

In addition to the static display of a community's attributes and properties, a further interesting visualization and analysis feature is the application of a time-related filter to the dataset of electronic communication. Adding new communication acts to the visualization algorithms according to their temporal occurrence allows observing, how the network actually formed over time (compare Figure 3). The observer can see how clusters are emerging and connecting with other clusters, or who has been the initial nucleus of a cluster. Here, a smooth and organic movement of nodes and visual growth of the graph is a major design issue.

Another useful analysis means developed in the project allows to filter the complex structures of the overall network (often involving thousands of authors) to reveal hidden core structures. This can be achieved by selecting a limited set of (important) authors or, as Figure 4 shows, a threshold of minimum relationship strength. This filter uncovers (compare left and right hand side) the backbone of a virtual communication network resulting from 10 days (02/04/2005 until 02/13/2005) of discussion activity in the public Java developer forum comp.java.lang.help.

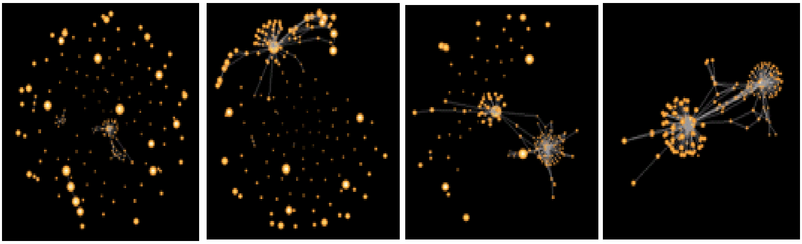

Figure 3: Screenshots of the evolution of an electronic communication network of a Corporate Instant Messaging Discourse. From initially unconnected employees, two clusters and a stable connection between them emerge

For moderators and coordinators, this helps to understand how the network works and how core contributors are embedded in their strong relationships. These indicate the authors' potential for utilizing social capital (compare next section) in a company.

Next to visualizations like showing the areas and clusters of dense relationships or high activity, highlighting the important authors or relationships and their evolvement over time, or revealing the core structures of complex networks, the software approach allows for eliciting ego-networks of different depths. The integration of keyword analysis enables to show the main keywords of the discourse, e.g. as an indicator of the competences of nodes. Further this allows for searching terms and reducing the network to only show the sub-networks which match those topics.

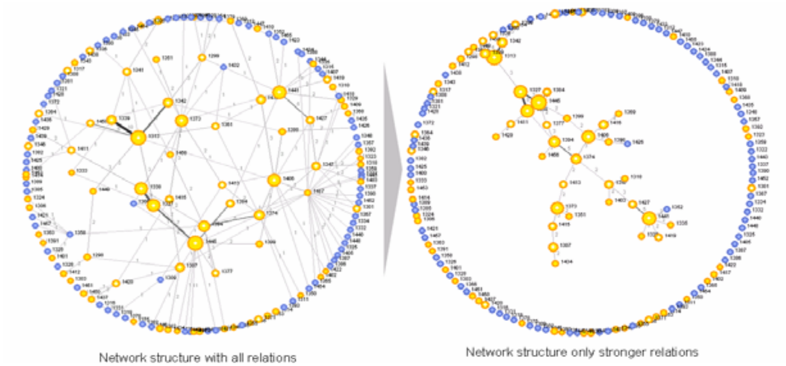

Figure 4: Breaking down complex networks to identify the core structures

Most of these visualizations already imply comparative measurements to give more quantitative accounts of the community structure under research, i.e. measures for density or activity. However, to explicitly generate absolute measures which enable the comparison between different discourses and thus aid the moderating tasks of communities, in the Commetrix project, a systematic system of measures for evaluating knowledge community networks is continually being examined and developed [also cf. Trie05c].

6. Evaluation Domains of Knowledge Networks

Schoen [Scho00] identified a selection of critical success factors, including aspects like user satisfaction, structures, contents and context, knowledge carriers, culture, and behavioral norms. From this comprehensive list, the following model with a more aggregated set of domains which can subsequently be evaluated using actual metrics and indicators from social network intelligence has been derived (Figure 5). It constitutes the methodical foundation for the evaluation and measurement approach for community networks that is continually being extended during the development of the Commetrix tool.

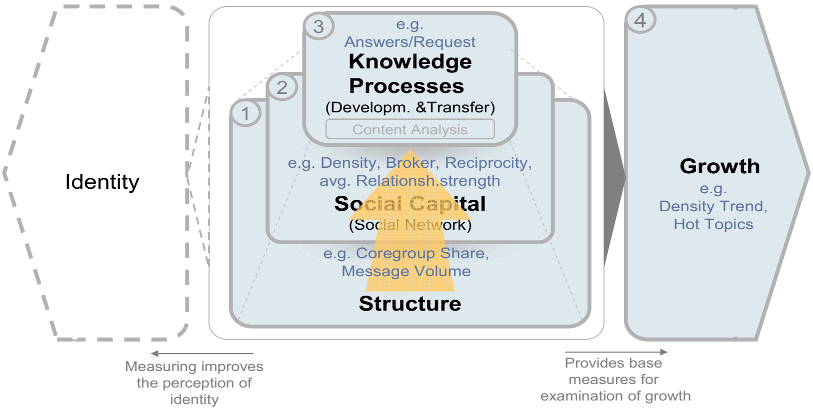

Figure 5: Towards a systematic measurement approach to evaluate Knowledge Networks - Layers and Sample Measures

The measurement concept for analyzing communication networks consists of four main domains (and one additional implicit output). The bottom layer (cf. Figure 5) is the actual quantitative structure. It can be evaluated by looking at measures like number of authors, volume of messages, average time between messages sent, network density, network diameter, etc. The structural measurements can also be calculated for each relation or for each author. An example for a more complex indicator of that domain is structural uniformity. It represents the evenness of the network in terms of activity and can have steep peaks (clear epicenters) or very little variance. Some structural properties simultaneously also include information about the social network within the communication network and by this also about the quality of social capital as the access to resources distributed in a people network [NaGo98]. Simple measures of this second domain are average strength of relationships, reciprocity of relations, number of direct and indirect contacts per author etc. After being able to capture such individual measures, an aggregated indicator could be derived to constitute a proxy for the level of trust.

Some of the structural properties additionally give insights about the level of knowledge exchange and its underlying knowledge processes. Offered measures of this third domain are unanswered versus answered questions, average number of replies, or average message length per relation. This knowledge related set of measurable network properties also leads to content-oriented analysis of the network structure. In the Social Network Intelligence approach of the project introduced in this paper, at this point automated keyword retrieval is being combined with social network analysis in order to align the analysis of communication networks with the idea of knowledge and expertise management. Resulting network indicators ('measures') are for example the main keywords of a relation, an author, or the complete network as well as similarity between authors based on keyword overlap.

These structural properties also form the basis for the final measurement domain: longitudinal growth and development. It helps to understand not only the configuration of the network, but also its dynamic behavior. This includes the identification of the group's collaboration velocity or deceleration, its declining sectors or the observation of the establishment of network roles over time. As already shown in the section on visualization algorithms, a longitudinal analysis component of the network has been developed for the software application. It allows to actually observes how new authors join in and form relations over time and how network properties are changing. In the future, this element will be developed and tested further to provide analytical insights about longitudinal network evolvement (cf. Figure 4).

Although it does not represent an actual domain of measurement, one final element of the model in Figure 5 represents the tight connection of measurements and visualization: the identity of the community. The group's identity does not easily lend itself to quantitative expression but is to some extend represented in the visual network structures, its coded colors and other graphical elements. This aspect is regarded as an implicit output of network analysis with high importance, as facilitation of virtual communication networks can also be spurred by simply visualizing information about otherwise invisible group structures. Examples are cues of who is similar to an ego, who is adjacent or co-present. This aspect has been identified and termed Social Translucence by Erickson and Kellogg [ErKe00].

7. Conclusion and Future Work

Visualization and evaluation of virtual communication networks can yield much insight about the domain of virtual knowledge work in Communities of Practice. The currently reduced transparency of these work structures can result in foregoing many of their benefits. Communities are being established but not nurtured using IT support, although the most data needed to create the required transparency exists in communication archives. The Commetrix project recognizes this challenge by exploring and researching innovative ways and means to utilize the rich set of communication data traces in order to help members, moderators, and researchers to better understand the invisible complex processes and structures of informal communication. This can eventually be employed in corporations to analyze the effectiveness of informal expert networks.

From a researcher's perspective, the set of visualizations is currently constantly being improved to create the best possible insights. In the measurement domain the set of implemented quantitative indicators will be extended to allow for better comparison of different discourses. The set of time-related dynamism measures like velocity or deceleration is of special interest, as no such network measures methodically exist by now. A last future challenge is the improvement of current text analysis and its synergetic integration into the social matrices in order to develop a potent methodology and the according tool for Social Network Intelligence for network-oriented knowledge work in virtual Communities of Practice.

References

[AmCo04] Ambrozek, J., Cothrel, J., 2004, Online Communities in Business: Past Progress, Future Directions. 7th International Conference on Virtual Communities, July 2004

[APQC03] APQC, 2003, Assessing the Health and Effectiveness of Communities of Practice. http://old.apqc.org/free/ articles/dispArticle.cfm?ProductID=1389

[Baec99] Baecker, D., 1999, Organisation als System (in German). Frankfurt am Main: Suhrkamp

[Barl96] Barley, S. R., 1996, Technicians in the workplace: Ethnographic evidence for bringing work into organization studies. Administrative Science Quarterly 41(3), pp. 404-441

[BeCo00] Berge, Z.L., Collins, M.P., 2000, Perceptions of e-moderators about their roles and functions in moderating electronic mailing lists. Distance Education: An International Journal, 21(1), pp. 81-100

[ErKe00] Erickson, T., Kellogg, W.A., 2000, Social Translucence: An Approach to Designing Systems That Support Social Processes. ACMTransactions on Computer-Human Interaction 7(2000)1, 59–83

[FrRe91] Fruchterman, T. M. J., Reingold, E. M., 1991, Graph Drawing by Force-Directed Placement, Software - Practice & Experience. 11(1991)21, 1129-1164

[Krac91] Krackhardt, D., 1991, The strength of strong ties: The importance of philos in organizations. In N. Nohira and R. Eccles (eds.): Organizations and networks: Theory and practice. Cambridge, MA: Cambridge University Press, pp. 216-239

[LeSt01] Lesser, E. L., Storck, J., 2001, Communities of practice and organizational performance. IBM Systems Journal 40 (4), p.831-841

[NaGo98] Nahapiet, J., Ghoshal, S., 1998, Social Capital, Intellectual Capital and the organizational Advantage. Academy of Management Review April 1998, 243.

[Rhei93] Rheingold, H., 1993, The Virtual Community - Homesteading On The Electronic Frontier. New York

[Scho00] Schoen, S., 2000, Gestaltung und Unterstützung von Communities of Practice (in German). Dissertation, TU Munich

[StLi00] Stamps, J., Lipnack, J., 2000, A Systems Science of Networked Organisations. Proceedings of the World Congress on Systems Sciences ISSS 2000. URL: www.virtualteams.com/library/whpapers/ISSS_2000.pdf. Accessed: 2004-10-01

[Swan01] Swan, J., 2001, Knowledge Management in Action: Integrating Knowledge across Communities. Proceedings of IEEE Hawaii International Conference of Systems Sciences, Hawaii 2001

[Trie05a] Trier, M., 2005a, Commetrix - Project Overview, http://www.commetrix.de

[Trie05b] Trier, M., 2005b, IT-supported Visualization of Knowledge Community Structures. Proceedings of 38th IEEE Hawaii International Conference of Systems Sciences HICCS38, Hawaii, USA, Jan 2005

[Trie05c] Trier, M., 2005c, IT-supported Visualization and Evaluation of Virtual Knowledge Communities, Dissertation Thesis, Technical University Berlin, 2005, http://opus.kobv.de/tuberlin/volltexte/2005/ 1072/